R Twitter Analysis

Tassie Gniady

Cyberinfrastructure for Digital Humanities Manager
UITS Research Technology, Indiana University
Pervasive Technology Institute

October 27, 2017

Gniady, Tassie. ”R Twitter Analysis." 27 Oct 2017. Digital Tools and Visualization Methods Series. Scholars' Commons, Wells Library, Indiana University, Bloomington. Retrieved from: http://hdl.handle.net/2022/21787

Digital Arts and Humanities Workshop Series – Fall 2017

Fridays @ noon -- Scholars Commons IQ-Wall

Date	Topic	Presenter
Aug. 25	Intro to Visualization	Michael Boyles
Sep. 1	Intro to Digital Humanities	Tassie Gniady
Sep. 8	Virtual Reality	Bill Sherman
Sep. 15	Intro to R	Tassie Gniady
Sep. 22	Advanced Media	Chris Eller
Sep. 29	Augmented Reality	Chauncey Frend
Oct. 13	R for Text	Tassie Gniady
Oct. 20	Network Graphs	David Kloster
Oct. 27	R for Twitter	Tassie Gniady
Nov. 3	3D Scanning and Printing	Jeff Rogers
Nov. 10	3D Photogrammetry	Tassie Gniady
Dec. 1	IQ-Tables & Touch-Enabled Software Workflows	David Reagan

Follow along in today's presentation for helpful links:

https://iu.box.com/v/rtwitter

What is Sentiment Analysis?

- Sentiment analysis is also known as **opinion mining** — it is used to extract subjective information from text.
- The word "subjective" is exactly why sentiment analysis can be difficult
- "I love chocolate ice cream as much as I love the sound of nails on a chalkboard."
  - "I hated nothing about that movie! Nothing!"
- Let's not even get started with emojis 😎🤔💩

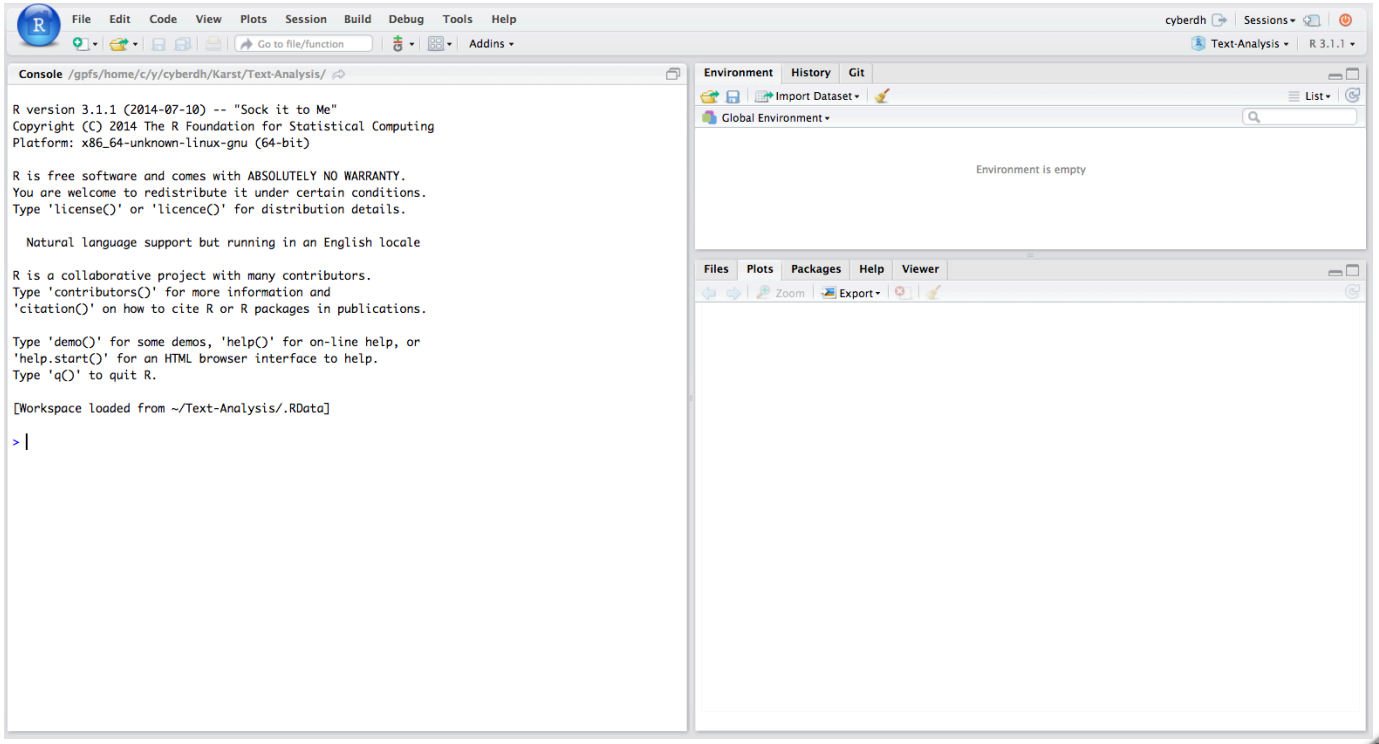
Karst Account

If you don't already have a Karst account, go ahead and request one right now.

https://kb.iu.edu/d/bezu#account

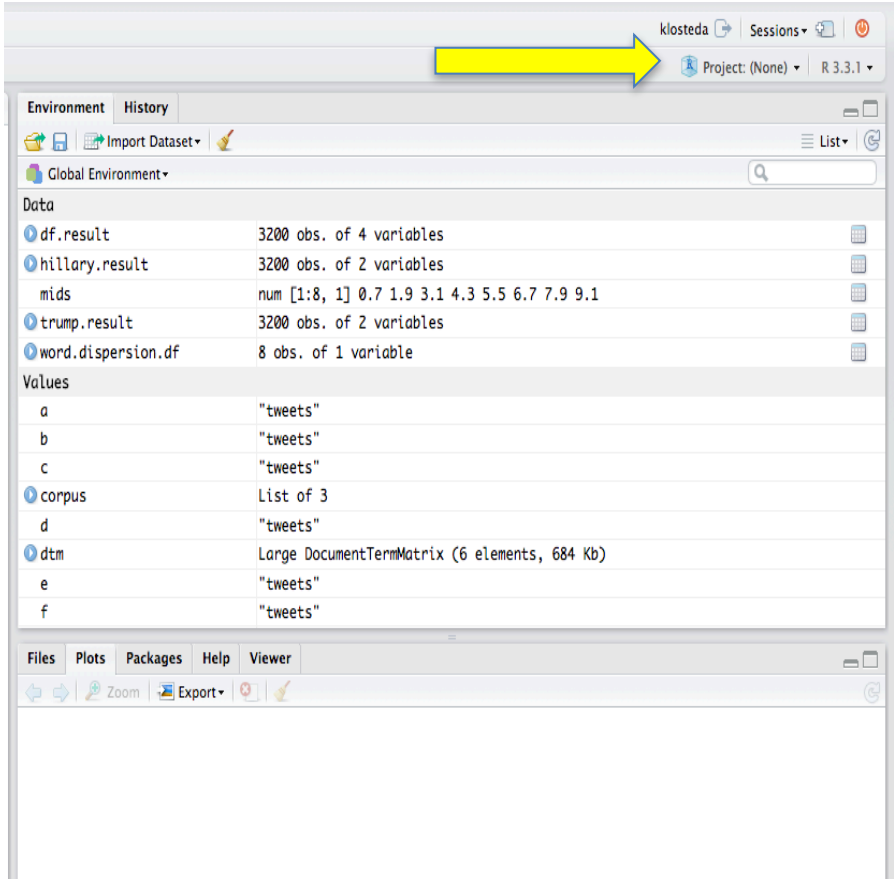
RStudio on Karst

- Go to https://rstudio.iu.edu

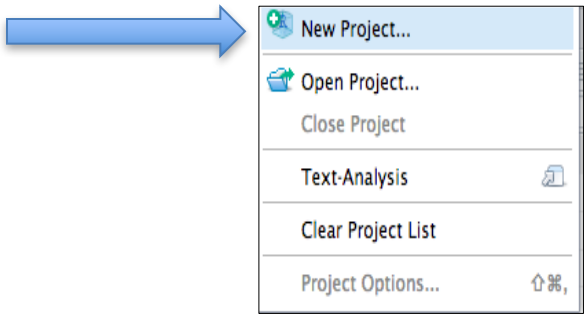


Creating a Project in R Studio from GitHub

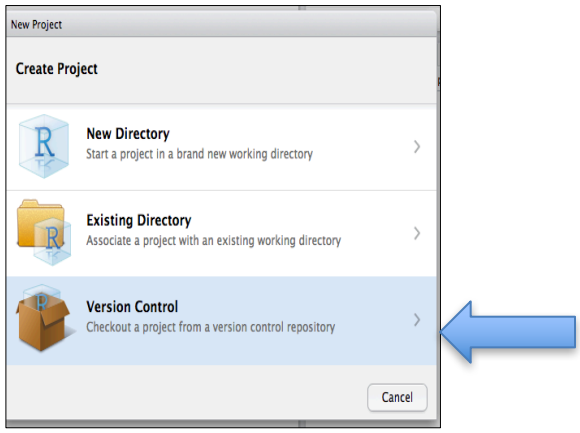
In R Studio, click on drop down arrow that says "Project: (None)"



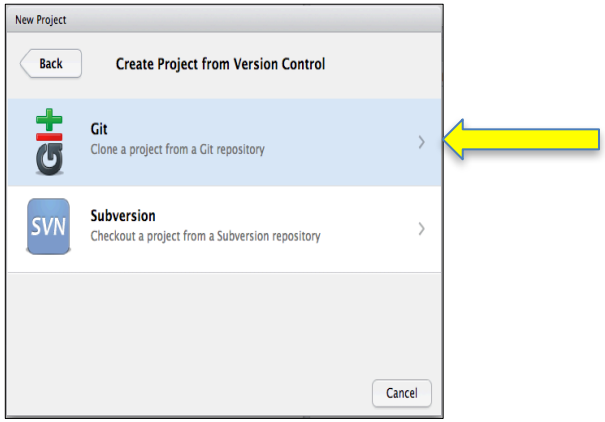
- Click on the first option "New Project…"



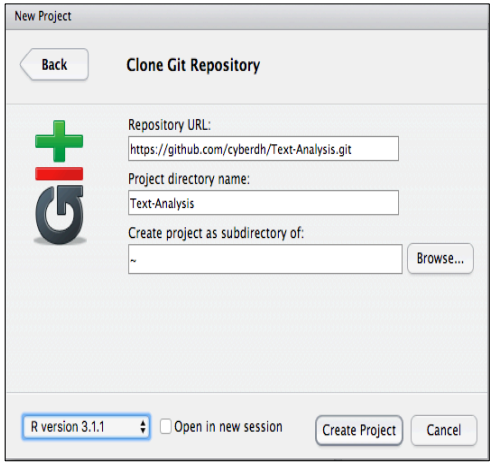
- A box should now appear asking you to choose between "New Directory," "Existing Directory," and "Version Control." Choose "Version Control."



- Another box should appear that asks you to choose between “Git” or “Subversion (SVN).” Choose Git.

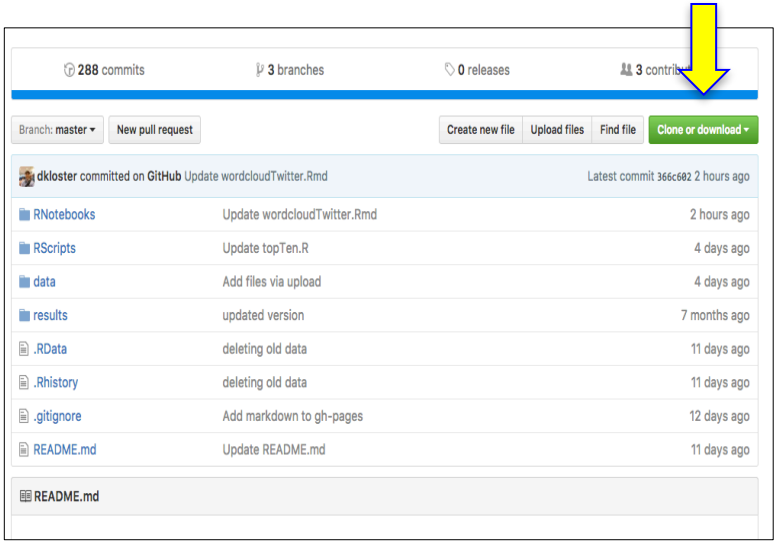


- This box should now appear asking for a “Repository URL” and directory name. For the URL we will go to the Cyber DH Text-Analysis GitHub repository at https://github.com/cyberdh/Text-Analysis.git

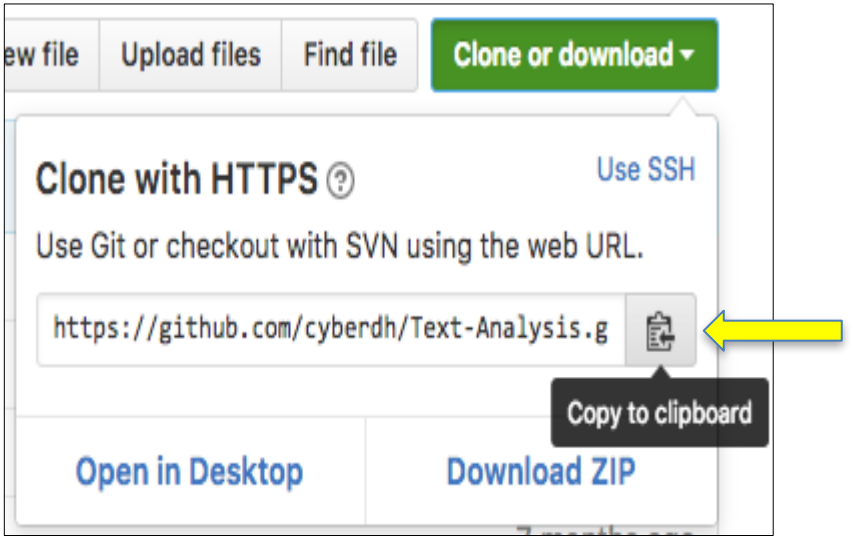


(make sure you open the page in another tab)

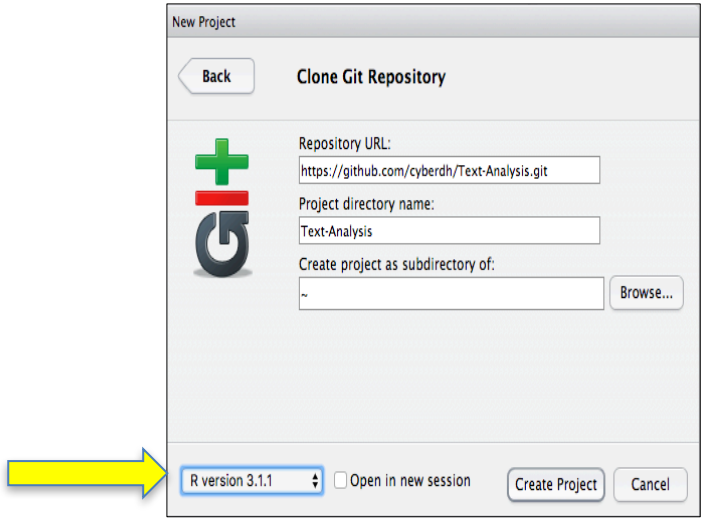
- The Cyber DH Text-Analysis repository on GitHub should look like this. Click on the green button labeled "Clone or download."

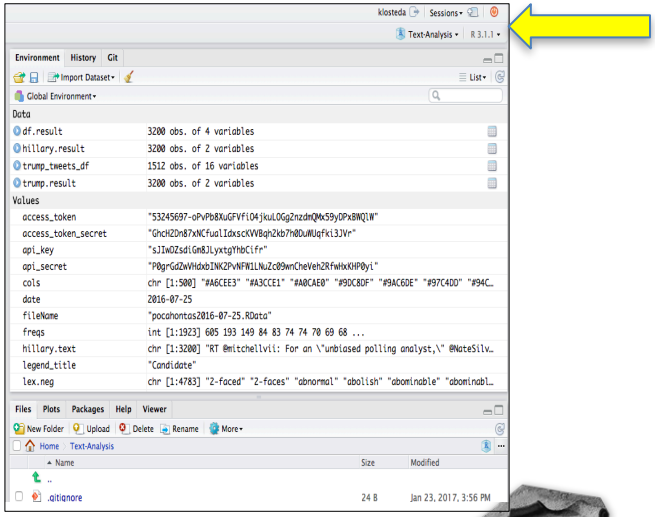


- This dropdown menu should appear. Make sure it says "Clone with HTTPS" in the top left corner. Then click the button with the clipboard at which the arrow is pointing. This will save it to your clipboard. Now go back to the R Studio page.



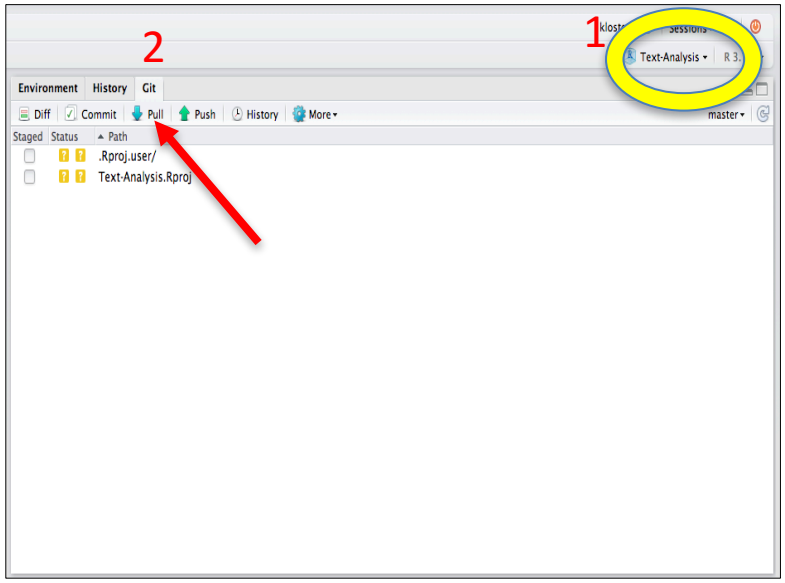
- Now paste the copied URL in the "Repository URL:" space. Next make sure the "Project directory name:" is **Text-Analysis** exactly including capitalization and dash. Otherwise you will need to change this in the script every time. Also, make sure you have R version 3.1.1 chosen in the lower left corner where the blue arrow is pointing. Then click the "Create Project" button.
- You should now see "Text-Analysis" listed in the project box and "R 3.1.1" in the box just to the right. You now have the Text-Analysis repository loaded on Karst.

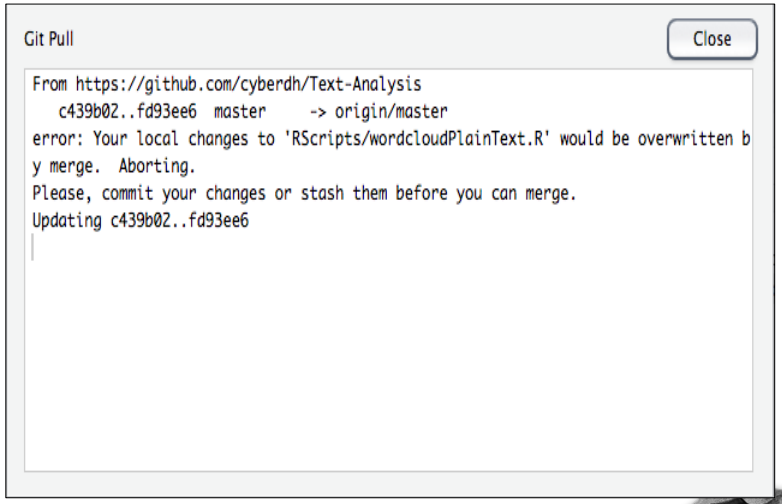




Pulling updates from Cyber DH GitHub repository on R Studio (Karst or Desktop)

1. Make sure you are in the Text-Analysis project. It should say it in the upper right hand corner.
2. Next, make sure you are in the **Git** tab. Now, there should be a button with a blue arrow pointing down that says **Pull**. Click on that button.
3. As you make changes, you will have to "stash" them locally.





Source Material

This script was adapted from Jeffrey Breen's tutorial "Mining Twitter for Airline Consumer Sentiment." Breen developed this is simple algorithm to show how data can be easily extracted and quickly synthesized.

http://www.inside-r.org/howto/mining-twitter-airline-consumer-sentiment

Overview of process

1. Get data (in this case Twitter has been mined for you and saved into .Rdata files)
2. Load and inspect data
3. Extract text
4. Load opinion lexicons
5. Implement algorithm
    1. Test algorithm with sample data first
6. Score twitter data using algorithm
7. Graph the result of scoring

Get data

- We used TwitteR
- cruz.tweets = searchTwitter('@tedcruz', n=4000, since="2016-02-02", until="2016-02-03")
- trump.tweets = searchTwitter('@realDonaldTrump', n=4000, since="2016-02-02", until="2016-02-03")
- clinton.tweets = searchTwitter('@HillaryClinton', n=4000, since="2016-02-02", until="2016-02-03")
- sanders.tweets = searchTwitter('@BernieSanders', n=4000, since="2016-02-02", until="2016-02-03")
- First tweets were gathered at 2016-02-02 at 11:59pm GMT going back 4,000 from that time

Load Opinion Lexicons

- We need lists of positive and negative words to use as a guide for scoring:

Liu, Minqing Hu and Junsheng Cheng. "Opinion Observer: Analyzing and Comparing Opinions on the Web." Proceedings of the 14th International World Wide Web conference (WWW-2005), May 10-14, 2005, Chiba, Japan. http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

- Can add words using the combine c() function to ensure the lists have words relevant to your corpus

Test the Algorithm with Sample Data

- We have written a function to compile and analyze the tweets (more on that in a minute).
- Before running the algorithm on all 4,000 tweets, lets test it first with these sentences:
- "This ice cream is the best! I love this flavor!"
- "I am so angry at the terrible weather!"
- "Impressed and amazed: you are peerless in your achievement of unparalleled mediocrity."

Implement Scoring Algorithm

- Let's write a function!
- One of the great strengths of R is the user's ability to add functions. In fact, many of the functions in R are actually functions of functions.
- Let's look again at the RNotebook

Going Further #zombiefrappuccino!

Use the wordcloud script from last week to look at the top words used to get more insight.

More Information

- Reference this presentation at: https://iu.box.com/v/rtwitter
- Get to know the Text Analysis repo through our GitHub Page Text Analysis Toolkit
- If you'd like to make an appointment, please email cyberdh@iu.edu.
- Also, check out our website: https://www.indiana.edu/~cyberdh/ for interesting blog posts and written tutorials about R.

Contact:
Tassie Gniady
Cyberinfrastructure for Digital Humanities Manager
ctgniady@iu.edu
@tassietheg